Densito (verde)

CAP I: INTRODUCCION A LA TEORIA DEL MUESTREO.

- CENSO: Información de todas las unid. de la población.
- UNIDADES DE MUESTREO: Elementos s/ los q. se realiza el muestreo.
  - ← Conjunto de: MARCO.
- FRACCION DE MUESTREO: Porcentaje de la población encuestado.

PROCESO GRAL. OBTENCION DE UNA MUESTRA.
- 1º Determinación estadística de la población.
- 2º Selección y tamaño de la muestra.
- 3º Preparación material impreso de la encuesta.
- 4º Organización del trabajo de campo.
- 5º Tratamiento estadístico de los rdos. de la encuesta.

ERRORES EN EL MUESTREO: a) Error no muestral o sistemático
- ERROR DE COBERTURA (pbción objeto ≠ pbción marco).
- FALTA DE RESPUESTA (↓ precisión , sesgo en rdo.).
- INSTRUMENTOS MEDICION INCORRECTOS, CUESTIONARIOS MAL DEFINIDOS, TECNICAS ENTREVISTA INADECUADAS.

b) Errores muestrales (consecuencia del uso de una muestra).

- DIFERENCIA entre ESTADÍSTICO MUESTRAL y PARÁMETRO POBLACIONAL.

* Este error se comete siempre.

- OBJETIVO FUNDAMENTAL en el DISEÑO DE UNA MUESTRA → MINIMIZAR errores muestrales y no muestrales.

MEDICIÓN INDIVIDUAL = VALOR VERDADERO ± ERROR MUESTREO ± ERROR SISTEMÁTICO

- TAMAÑO DE LA MUESTRA: Nº de elementos de la población q. constituyen la muestra, depende de el NIVEL DE PRECISIÓN.

- VARIANZA POBLACIONAL: Cuanto + homogénea es la población objeto → VARIANZA < (nº elementos en muestra <).
  " − " " " → " > (" " " " >).

- NIVEL DE CONFIANZA: Elección subjetiva del investigador. Es la PROBABILIDAD de ACERTAR.

TIPOS DE MUESTREO.

OBJETIVO FINAL: Selección de una muestra q., al mín. coste, nos permita obtener la mejor representación de la PBCION.

a) MUESTREO NO PROBABILÍSTICO (minimizar costes).

a.1. MUESTREO DE CONVENIENCIA: muestra se elige por pura conveniencia.

a.2. " " JUICIOS: experto selecciona muestra q. a su juicio mejor representa a la pbción.

a.3. " " POR CUOTAS: en combinación con otros tipos de muestreo. El entrevistador, dada una cuota preestablecida de observaciones a obtener, elige los elementos a encuestar.

b) MUESTREO PROBABILÍSTICO (selección de la muestra al azar, probabilidad q. tiene c/ observación de la pbción de entrar en la muestra: conocida).

* Conclusiones generalizadas a la pbción. acompañadas de un ERROR MÁXIMO.

b.1. MUESTREO ALEATORIO SIMPLE:

(-) C/ elemento de pbción misma probab. de ser seleccionado para entrar en muestra. ("1/n" -> pbción: "n").

(-) Uso sólo si comportamiento HOMOGÉNEO, SIMILAR y UNIFORME.

b.2. MUESTREO ALEATORIO SISTEMÁTICO: (caso particular del anterior).

- Dividir el marco en tantos bloques iguales como observaciones hayan en la muestra.
- Seleccionar un nº al azar en el 1er bloque
- Selección de la muestra total: tomando en todos los bloques el elemento correspondiente en el orden al del seleccionado en el 1er bloque.

* Probabilidad elección misma para todos lo elementos de la pblación (n/N).

VENTAJAS: - sencillez y comodidad.
- muestra distribuida a lo largo del marco.

DESVENTAJAS: - posibilidad de sesgos.

b.3. MUESTREO ALEATORIO ESTRATIFICADO:

→ HOMOGÉNEOS en sí, HETEROGÉNEOS entre sí.

- Definición de la vble. o vbles. de estratificación.
- Construcción de los marcos de c/ estrato.
- Selección de una muestra en c/ estrato.

* PROBLEMA: Distribución del tamaño total de la muestra:

I. PROPORCIONALMENTE → entre los estratos en función del tamaño de c/ estrato con respecto a la pblción.

II. ÓPTIMAMENTE → en función de la dispersión de las observaciones contenidas en c/ estrato.

b.4. Muestreo aleatorio por conglomerados o clusters:

→ **Heterogéneos dentro de sí, Homogéneos entre sí.**

- ⊖ Construcción de conglomerados
- ⊖ Selección de la muestra de conglomerados. (al azar).

CAP. II: DISTRIBUCIONES MUESTRALES DE PROBABILIDAD.

- 4 PROPIEDADES DE UN BUEN ESTIMADOR:

a) Ser INSESGADO (su valor esperado coincide con el verdadero valor del parámetro estimado).

b) Ser EFICIENTE (prop. relativa: un estimador es + eficiente q. otro cuando tiene < varianza).

c) Ser SUFICIENTE (incluye toda la información q. la muestra puede proporcionar s/ el parámetro poblacional).

d) Ser CONSISTENTE (proporciona informaciones q. convergen en probabilidad hacia el parámetro poblacional).

- POBLACIÓN: Conjunto de N observacio. q. representan el "estado de naturaleza" de una carac. medible X.
- MUESTRA: Conjunto de n observaciones extraídos de una población de tamaño N ($n < N$).

NATURALEZA DE LA COMBINACIÓN LINEAL (Y):

- Si $X_i \Rightarrow$ NORMAL, entonces $Y \Rightarrow$ NORMAL, $\forall n$.
- Si $X_i \neq$ NORMAL, entonces $Y \Rightarrow$ NORMAL, si $n \to \infty$

⊛ DISTRIBUCIÓN MUESTRAL → "V.A. constituida por el valor de un estadístico calculado en todas las posibles muestras de tamaño "n" que se pueden extraer de la pbción. de partida."

Distribución Muestral de la Media:

Pbción de partida: Distr. $N(\mu_x ; \sigma_x^2)$.

Desviación estándar = error estándar.

* $\bar{X} = \frac{(x_1 + x_2 + \ldots + x_n)}{n}$; $E[\bar{X}] = \mu_{\bar{x}} = \mu_x$ ⇐ Siempre.

* $Var(\bar{x}) = \sigma_x^2 / n$; Si M.S.R. ⇒ $Var(\bar{x}) = \sigma_x^2 / n \left(\frac{N-n}{N-1}\right)$ → CPF → 1 cuando $N \to \infty$

Distribución Muestral de la Varianza:

Pbción de partida: Distr. $N(\mu_x ; \sigma_x)$.

* $S^2 = \frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}$ (cuasivarianza muestral)

* $E(S^2) = \sigma^2$

* $\chi^2_{n-1} = \frac{(n-1) S^2}{\sigma^2}$

* $Var.(S^2) = 2\sigma^4 / n-1$

* Nivel de confianza: área central subjetiva/ elegida.
* " " significación = 100% - nivel de confianza.

Distribución Muestral de la Media: Caso varianza pob. desconocida.

Pto. de partida:

$X \sim N(\mu_x; \sigma_x^2)$.

$\bar{X} \sim N(\mu_{\bar{x}} = \mu_x; \sigma_{\bar{x}}^2 = \sigma_x^2/n)$

$$Z = \frac{\bar{x} - \mu_x}{\sigma_x/\sqrt{n}}$$

Al desconocer el valor de la varianza poblacional, la sustituimos por la varianza muestral.

* La tipificación pasa a ser una "t" de Student:

$$t_{n-1} = \frac{\bar{x} - \mu_x}{s/\sqrt{n}}$$

; Si M.S.R. ⟹

$$t_{n-1} = \frac{\bar{x} - \mu_x}{\frac{s}{\sqrt{n}}\sqrt{1 - \frac{n}{N}}}$$

Distribución Muestral de la Varianza: Dos Poblaciones.

Pto. de partida: sacamos de dos poblaciones dos muestras de tamaño "n":

$X \rightarrow N(\mu_x; \sigma_x^2)$

$$\frac{(n-1) S_x^2}{\sigma_x^2} = \chi^2_{n_x-1}$$

$Y \rightarrow N(\mu_y; \sigma_y^2)$

$$\frac{(n-1) S_y^2}{\sigma_y^2} = \chi^2_{n_y-1}$$

$$\frac{\chi^2_{n_x-1}/(n_x-1)}{\chi^2_{n_y-1}/(n_y-1)} = F_{n_x-1;\, n_y-1} = \frac{S_x^2/\sigma_x^2}{S_y^2/\sigma_y^2}$$

Prueba de Homocedasticidad.

Si $\sigma_x^2 = \sigma_y^2 \Rightarrow \frac{S_x^2}{S_y^2} = F_{n_x-1;\, n_y-1}$.

DISTRIBUCIÓN MUESTRAL DE LA MEDIA: DOS POBLACIONES.

Pto. de partida:

$X \to N(\mu_x ; \sigma_x^2)$
$\downarrow$
$\bar{X} \to N(\mu_x ; \sigma_x^2/n_x)$

$Y \to N(\mu_y ; \sigma_y^2)$
$\downarrow$
$\bar{Y} \to N(\mu_y ; \sigma_y^2/n_y)$

$$\bar{X} - \bar{Y} \to N\left(\mu_x - \mu_y ; \sigma_x^2/n_x + \sigma_y^2/n_y\right)$$

1er CASO: VARIANZAS POB. CONOCIDAS PERO $\neq$:

$$z = \bar{x} - \bar{y} - (\mu_x - \mu_y) \Big/ \sqrt{\frac{\sigma_x^2}{n_x} + \frac{\sigma_y^2}{n_y}}$$

2º CASO: VARIANZAS POB. CONOCIDAS E IGUALES:

$$z = \bar{x} - \bar{y} - (\mu_x - \mu_y) \Big/ \sqrt{\frac{\sigma^2}{n_x} + \frac{\sigma^2}{n_y}}$$

3er CASO: VARIANZAS POB. DESCONOCIDAS PERO NO SIGNIFICATIVA/ $\neq$ (HOMOCEDASTICIDAD):

$$t_{n_x+n_y-2} = \frac{\bar{x} - \bar{y} - (\mu_x - \mu_y)}{\sqrt{\frac{S_p^2}{n_x} + \frac{S_p^2}{n_y}}}$$

$$* \quad S_p^2 = \frac{(n_x - 1) S_x^2 + (n_y - 1) S_y^2}{n_x + n_y - 2}$$

4º CASO: VARIANZAS POB. DESCONOCIDAS PERO $\neq$ (HOMOCEDASTICIDAD):

$$t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{\sqrt{\frac{S_x^2}{n_x} + \frac{S_y^2}{n_y}}}$$

* Si los grados de libertad son suficiente/ grandes, $t \to z$.

b) **M.C.R.: naturaleza BINOMIAL**

aproximación a NORMAL si $nPQ \geq 3$.

DISTRIBUCION MUESTRAL DE LA PROPORCION DE EXITOS:

$$\bar{p} = \frac{r}{n} = \frac{\text{Nº de éxitos}}{\text{Tamaño muestra}}$$

$$E(\bar{p}) = P \quad ; \quad Var(\bar{p}) = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$

$$Nat.(\bar{p}) = Nat(r)$$

HIPERGEOMÉTRICA → (N y/o n ≫) → BINOMIAL → ($nPQ \geq 3$) → NORMAL (corrector de continuidad: $1/2n$).

BINOMIAL → ($n \to \infty$, $P \to 0$) → POISSON

* DISTRIBUCIONES MUESTRALES DERIVADAS DE LA BINOMIAL:

- Vble. a estudiar de naturaleza CUALITATIVA. (ÉXITO o FRACASO).

DISTRIBUCIÓN MUESTRAL DEL NÚMERO DE ÉXITOS:

De la población de partida conocemos:

"P": Porcentaje o probabilidad de aparición de la modalidad éxito en la población.

"Q": " " " " " fracaso " " ": $(1-P)$

→ De la población extraemos todas las muestras posibles de tamaño "n". En cada una podemos calcular el estadístico "r" (nº de éxitos en la muestra):
los "k" valores de "r" constituyen una DISTRIBUCIÓN MUESTRAL DE ÉXITOS.

a) M.S.R.: naturaleza HIPERGEOMÉTRICA

$E[r] = nP$; $Var(r) = nPQ\left(\frac{N-n}{N-1}\right)$

* Probabilidad de que en una muestra de tamaño "n" extraída de una pob. de tamaño N, r tome un determinado valor k:

$$P[r = k] = \frac{\binom{NP}{k}\binom{NQ}{n-k}}{\binom{N}{n}}$$

* Si N y/o >> ⇒ MODELO HIPERGEOMÉTRICO
↓ a
MODELO BINOMIAL.

$nPQ \geq 3$
* corrección de continuidad